

EINE ONTOLOGIE ZUR DEUTSCHEN GRAMMATIK

von Ineta Sejane und Roman Schneider

Im Rahmen des Projekts „Texttechnologie für die Grammatik“ wird am Institut für Deutsche Sprache derzeit eine domänenspezifische Fachontologie zur deutschen Grammatik entwickelt. Um einen ersten Einblick in die damit verbundenen Arbeitsschritte und Forschungsziele zu ermöglichen, sollen nachfolgend einige forschungspraktische Überlegungen der Projektgruppe formuliert werden.

Die Idee, dass wir eine eigene, auf das grammatische Online-Informationsangebot des Instituts abgestimmte terminologische Basis benötigen, konkretisierte sich Schritt für Schritt im Verlauf einer Auswertung von Benutzerwünschen des grammatischen Hypermedia-Informationssystems **grammis**.¹ Wiederholt meldeten hierbei interessierte Rezipienten einen Bedarf an verstärkter terminologischer Hilfestellung sowie feinkörnigeren Rechercheoptionen an. Nach umfangreichen konzeptuellen Vorarbeiten konnte schließlich Anfang dieses Jahres die praktische Arbeit an der Umsetzung dieses Desiderats aufgenommen werden. Wir wollen mit der Ontologie nicht nur zur Konsolidierung der in **grammis** verwendeten grammatischen Fachsprache beitragen, sondern insbesondere auch zur Entwicklung konkreter Angebote hinter dem populären Schlagwort vom „Semantischen Netz“.² Ziel unserer Bemühungen ist es, Computerprogrammen einen ausgesuchten Teil unseres Weltwissens „beizubringen“, damit sie dem menschlichen Informationssuchen situations- und kontextabhängig effiziente Hilfestellung anbieten können.

Im Sprachgebrauch der Informationstechnologie bzw. des Information-Retrievals modelliert eine Ontologie das Wissen,

das wir über ein Fachgebiet (Domäne) besitzen.³ Was man dabei unter dem Begriff ‚Wissen‘ versteht, kann in Form und Umfang von Mensch zu Mensch – je nach Vorbildung und Interessen – sehr unterschiedlich ausgeprägt sein. Unser Benutzerkreis wird vor allem aus akademischem Fachpublikum bestehen, es sollen aber gleichermaßen auch interessierte Laien wie z.B. Schüler oder Fremdsprachenlerner einen adäquaten Zugang zu grammatischen Informationen finden.

Bei der Alternative, die Grammatik-Ontologie aus Fachtexten⁴ oder GermaNet bzw. ähnlichen semantisch-lexikalischen Ressourcen automatisch zu extrahieren oder sie manuell zu erstellen, haben wir uns für die überwiegend manuelle Erstellung entschieden. So wird das Wissen besser, weil kontrolliert modelliert, und zugleich können die Bedürfnisse der verschiedenen Benutzergruppen hinsichtlich der Terminologieverwendung eher berücksichtigt werden. Wir bedienen uns beim Erstellen der Ontologie allerdings auch statistischer, korpusbasierter Methoden; so haben wir beispielsweise ein algorithmisiertes Verfahren entwickelt, um Konzeptkandidaten (=Termini) aus einem bestehenden Fachtextkorpus abzuleiten. In sukzessiv

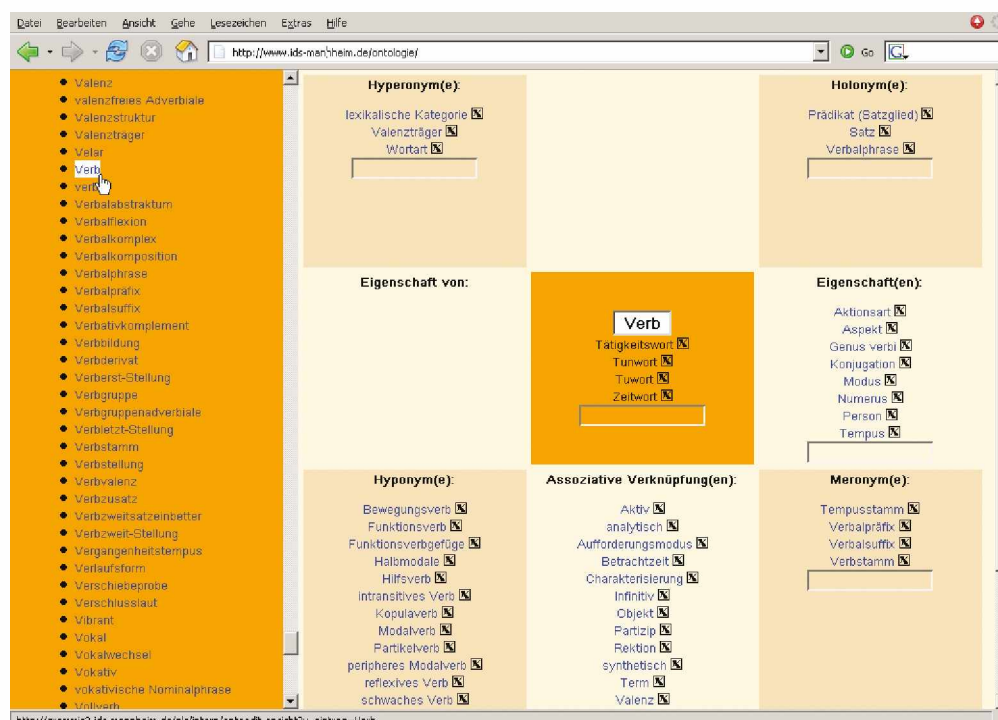


Abb. 1: Autorensicht der Ontologie

aufeinander aufbauenden Arbeitsschritten wird eine Kandidatenliste unter Berücksichtigung verschiedener Parameter (Frequenzanalyse des Fachtextkorpus sowie allgemeinsprachlicher Korpora, Kollokationsanalyse, Markup-Analyse) extrahiert und abschließend den Fachexperten zur Weiterverarbeitung zur Verfügung gestellt.

Das Online-Informationssystem **grammis** – wie bereits erwähnt sowohl die terminologische Grundlage als auch das praktische Anwendungsgebiet der Ontologie – beruht inhaltlich im Wesentlichen auf der dreibändigen „Grammatik der deutschen Sprache“⁵. Es umfasst derzeit fünf Komponenten, nämlich die „Systematische Grammatik“, das „Terminologische Wörterbuch“, das „Grammatische Wörterbuch“, die „Grammatische Bibliografie“ sowie das neue Modul mit dem Titel „Grammatik in Fragen und Antworten“, welches einen dezidiert problemorientierten Zugang ermöglicht.⁶ Alle diese Ressourcen sollen mit Hilfe der Ontologie verstärkt miteinander verknüpft werden und, was besonders wichtig ist, die Suche in den angeschlossenen Hypertexten soll flexibler und effizienter ausgestaltet werden.

Abbildung 1 vermittelt einen Eindruck der von uns zum Aufbau der Ontologie verwendeten Autorenumgebung. Im Zentrum des Bearbeitungsformulars befindet sich der aktuell aufgerufene Terminus. Um ihn herum gruppiert sind Eintragsfelder für unterschiedliche Relationstypen wie Synonymie, Hyperonymie/Hyponymie, Holonymie/Meronymie, (vererbare) Eigenschaften sowie assoziative Verknüpfungen (Kookkurrenzen). Jeder in der Ontologie eingetragene Terminus wird in der alphabetisch geordneten linken Spalte angezeigt, darüber hinaus existieren mehrere hierarchisch orientierte Darstellungsvarianten. Eine Navigation im System ist sowohl durch das direkte Auswählen von Listenelementen als auch durch das Anklicken der Relationspartner möglich. Zur Bearbeitungszeit überprüfen im Hintergrund arbeitende Datenbank-Trigger automatisch die Einhaltung grundlegender Kohärenzregeln: So darf ein Terminus beispielsweise nicht gleichzeitig als Hyperonym und Hyponym eines zweiten Terminus eingetragen werden; analog gilt dies im Falle von Teil-Ganzes-Beziehungen für Holonyme und Meronyme. Synonyme bilden einen Synonymring, den wir in konzeptueller Anlehnung an die populären Wort-Wissensnetze „WordNet“ und „GermaNet“⁷ auch „Synset“ nennen. Als Bezeichner eines Synonymrings wählen wir, wo immer möglich, denjenigen Terminus aus, der in der „Grammatik der deutschen

Sprache“ bzw. in **grammis** bevorzugt verwendet wird. Als Synonyme betrachten wir bewusst auch partielle Synonyme, die zum größten Teil verschiedenen terminologischen Systemen – also z.B. unterschiedlichen Grammatiken – entspringen; allerdings markieren wir entsprechende Fälle.⁸

In Abbildung 2 lässt sich die derzeit im Entwicklungsstadium befindliche Rechterschnittstelle erkennen. Konkret kann damit in der „Systematischen Grammatik“ sowie in der „Grammatischen Bibliografie“ nach Treffern gefahndet werden. Die dargestellten Termini lassen sich per Mausklick als Hyperlinks aktivieren und damit zur Navigation innerhalb der (Hierarchie bzw. Teilhierarchie der) Ontologiebasis nutzen. Die gezielte Suche erfolgt dadurch, dass der Benutzer einen Begriff aus der Konzeptfamilie (die um den Synonymring in der Mitte gruppierten Termini) auf der linken Seite des Bildschirms auswählt und ihn per Drag-and-Drop in einen der drei Container auf der rechten Seite zieht. Daraufhin wird automatisch der Umfang der betreffenden Ergebnismenge berechnet und numerisch dargestellt. Gleichmaßen verfährt man mit weiteren Suchbegriffen. Sowohl die Ergebnismenge für jeden einzelnen Suchbegriff als auch die Schnittmenge zwischen diesen Ergebnismengen werden angezeigt. Diese lassen sich ebenfalls unmittelbar als Hyperlinks aktivieren und zum Überwechseln in die „Systematische Grammatik“ bzw. die „Grammatische Bibliografie“ nutzen.

Die Grammatik-Ontologie soll nicht nur als eigenständige Applikation das gezielte Navigieren in **grammis** befördern. Sie ist auch als Hilfsmittel für die Volltextrecherche in bereits etablierten Komponenten konzipiert. Dabei lassen sich nicht nur einfache Suchwörter, sondern auch komplexere Suchabfragen eintippen. Bei Eingabe eines allein stehenden Terminus oder einer Phrase werden lediglich diejenigen Hypertexte gefunden, die exakt diesen Terminus bzw. diese Phrase enthalten. Gibt man aber den Operator SYN (synonym) wie z.B. „SYN(Verb)“ ein, so erweitert das System die Abfrage um die betreffenden, in der Ontologie hinterlegten Synonyme; im Falle von ‚Verb‘ also um ‚Tätigkeitswort‘, ‚Tuwort‘ und ‚Zeitwort‘. Weitere Operatoren sind BT (broader term) für Hyperonyme, NT (narrower term) für Hyponyme, BTP (broader term partial) für Holonyme, NTP (narrower term partial) für Meronyme und RT (related term) für assoziative Verknüpfungen bzw. Kookkurrenzen. NT(Valenz) findet also z.B. auch Hypertexte, in denen das Hyponym „Verbvalenz“ vorkommt. Solche Operationen dürfen

selbstverständlich mit den bekannten logischen Operatoren AND, OR und NOT zu beliebig komplexen Anfragen kombiniert werden. Einerseits erweitert sich durch die Einbindung der Ontologie der Trefferraum, andererseits kann durch exaktere Anfrageformulierung die Ergebnismenge auch stark reduziert werden, indem ungewollte Suchresultate automatisch ausgefiltert werden. Über das genaue Prozedere beim Anbinden der Ontologie an die Volltextrecherche soll noch entschieden werden. Neben der geschilderten Abfragesyntax ist auch eine benutzerfreundlichere Variante denkbar, die mit definierten Suchprofilen (z.B. „nur Synonyme hinzufügen“ oder „nur Hyperonyme hinzufügen“ etc.) arbeitet.

grund schrumpfender Bibliotheksetats sowie der Tatsache, dass – nicht zuletzt aufgrund der in den letzten Jahren konkretisierten politischen Publikationsbedingungen für öffentlich geförderte Forschungsprojekte – immer mehr Forschungsergebnisse online veröffentlicht werden, erscheint dieses Vorhaben von beträchtlicher Relevanz für die einschlägige wissenschaftliche Recherche.

Anmerkungen

¹ Das Angebot ist im Internet erreichbar unter www.ids-mannheim.de/grammis/.

² Vgl. hierzu www.w3.org/2001/sw/ oder www.semanticweb.org/.

³ Eine gängige Formulierung hierfür ist „Spezifikation einer Konzeptualisierung“; vgl. z.B. Schneider (2006).

⁴ Siehe z.B. Leitner (2006) oder den Sammelband „Ontology Learning from Text: Methods, Evaluation and Applications“ (2005).

⁵ Zifonun et al. (1997).

⁶ Für nähere Informationen vgl. Strecker (2005), Schneider (2004) und Frosch (2005) sowie die Beiträge von Konopka und Strecker in SPRACHREPORT 3/06.

⁷ Siehe <http://wordnet.princeton.edu> bzw. www.sfs.uni-tuebingen.de/lsd/.

⁸ Eine Modellierung, die gezielt mit der Problematik der unterschiedlichen Verwendung eines Terminus in verschiedenen grammatischen Schulen bzw. Theorien umgeht, ist für die Zukunft geplant.

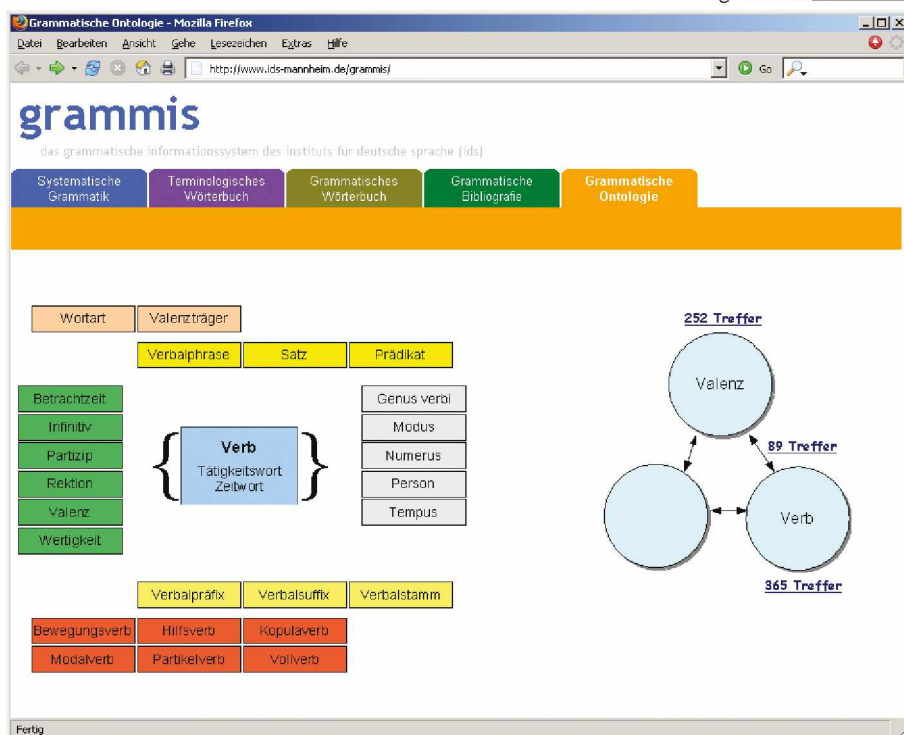


Abb. 2: Ontologie-basierte Volltextrecherche für **grammis** (in Entwicklung)

Insgesamt wollen wir mit der Ontologie den Benutzern von **grammis** ein Werkzeug an die Hand geben, welches bei der Klärung von Fragestellungen zur grammatischen Terminologie eine erste Orientierungshilfe bietet und auf einschlägige Online-Komponenten weiterverweist. Des Weiteren möchten wir den Benutzern vermittelt einer intuitiv gestalteten Schnittstelle variable semantische Rechercheoptionen eröffnen. Mittelfristig möchten wir die Ontologie auch für weiterführende Aufgaben einsetzen; interessant erscheinen etwa eine automatisierte Textklassifizierung und -verschlagnung sowie die Suche in verteilten Internetressourcen. In diesem Zusammenhang ist insbesondere an die wachsende Popularität sogenannter Open-Access-Plattformen zu denken: Vor dem Hinter-

Literatur

Berners-Lee, Tim / James Hendler / Ora Lassila (2001): The Semantic Web. In: Scientific American, May 2001. URL: www.scientificamerican.com.

Buitelaar, Paul / Philipp Cimiano / Bernardo Magnini (Hg.) (2005): Ontology Learning from Text: Methods, Evaluation and Applications. Amsterdam u.a.: IOS Press. (= Frontiers in Artificial Intelligence and Applications, Vol. 123)

Frosch, Helmut (2005): BDG – Die Bibliografie zur deutschen Grammatik. In: SPRACHREPORT 3/05, S.15-16.

Leitner, Johannes (2006): Extraktion von Ontologien aus natürlichsprachlichen Texten. URL: www.inf.uni-konstanz.de/~leitner/docs/ontomining.pdf.

Schneider, Roman (2006): Texttechnologie und Grammatik. In: Breindl, Eva / Lutz Gunkel / Bruno Strecker (Hg.)

- (2006): Grammatische Untersuchungen. Analysen und Reflexionen. Tübingen: Narr, S. 601-616.
- Schneider, Roman (2004): Benutzeradaptive Systeme im Internet: Informieren und Lernen mit GRAMMIS und ProGr@mm. Mannheim: IDS (=amades 4/04).
- Strecker, Bruno (2005): Grammis – Das grammatische Informationssystem des Instituts für Deutsche Sprache. In: SPRACHREPORT 3/05, S. 12-15.
- Zifonun, Gisela/Ludger Hoffmann/ Bruno Strecker (1997): Grammatik der deutschen Sprache. 3 Bände. Berlin/New York: de Gruyter. (= Schriften des Instituts für Deutsche Sprache 7, 1-3)
- Die Autoren sind wissenschaftliche Mitarbeiter am Institut für Deutsche Sprache in Mannheim.